Research on In-use Inspection Method of Ultrasonic Gas Flow Meter Based on Supervised Learning

C.Z. Lü[1], W.L. Li[2,3], M.N Li[2], C.H. Li[2*], D.L. Xie[3]

[1] National Pipeline Group Beijing Pipeline Company Limited, Da Tun Lu 9#, Beijing, China
[2] National Institute of Metrology, Beijing, China
[3] China Jiliang University, Hangzhou, Zhejiang, China
E-mail (corresponding author): lich@nim.ac.cn

Abstract

Ultrasonic flowmeters are the most commonly used instruments in natural gas trade, the in-use inspection methods can be used to verify the performance of it. The supervised learning algorithm with the digital measurement technology is utilized to investigate the method in-use of the ultrasonic flowmeter. Random Forest and BP-Artificial Neural Network are used to construct the soft-measurement models to estimate flow rate deviation. With the close loop gas flow standard facility of NIM in China, there were 6 flow points selected to conduct test experiments on the DN200 ultrasonic gas flowmeter with cross 4-paths. Together with the test data, 15 indicators are determined as the input of the model, and the data is denoised by means of Fast Fourier transform with Gaussian window function. MCM is used to assess the uncertainty. The results show that the two models can estimate well the flow rate deviation of the flowmeter, in which Random Forest model has the better result ($\overline{R^2} \geq 0.74$, $\overline{E}_{RMSE} \leq 0.06$ %, $u(y) \leq 0.12$ %), with the advantages of high accuracy and good stability, it can effectively monitor and diagnose the performance of flowmeter during the operation. Thus, a new reference is provided for the inspection process of ultrasonic flowmeter in use.

1. Introduction

Natural gas is a low-carbon, clean and environment-friendly energy source with increasing demand in the worldwide. The time difference ultrasonic gas flowmeter (USM), characterized by no moving parts, high measurement accuracy and good self-diagnosis function, is one of the commonly used measuring instruments in natural gas trade at present. In order to ensure the accuracy of measurement, the USM needs to be verified periodically in the verification agency with the flow standard facility, it also can be checking by in-use inspection method based on the verification results, so as to avoid potential safety hazards in the process of disassembly and assembly of the natural gas instrument, and thus improve the use efficiency of the instrument.

At present, speed of sound (SoS) checking is the most common method of USM in-use inspection from AGA No.10 Report [1] and ISO 20765-1 [2], in which the speed of sound deviation is based on the calculation of speed of sound with the thermodynamic state equation, it can reflect whether the flowmeter meets the performance requirements. Research on the inspection method of USM, in 2012, the German Federal Institute of Physical Technology (PTB) published the technical guideline G-18 proposing [3] : for the checking of two USMs installed in series, differences between meter indication (the measurement flow rate deviation) shall not change more than 0.5% compared to the initial situation, and the differences in SoS-indication among the paths (the sound of speed deviation) shall not change more than 0.3% compared to the initial situation. On this basis, in 2020 [4], the Nanjing station of Pipe-China explored that the method of inspection with the multivariate linear estimation based on the historical data from two flowmeters installed in series, which could monitor if the flow rate deviation of the flowmeter exceeds the dynamic warning limit. However, up to now, the relationship between the speed of sound deviation and measurement flow rate deviation was not provided.

In this paper, the soft-measurement model (estimation model of measurement flow rate deviation) based on supervised learning is presented, which combined existing research methods with digital metrology technology, it can connect all diagnostic parameters of instruments with measurement flow rate deviation by verification data under the condition of real flow with the aid of a flow standard facility, meanwhile, this process can also monitor and diagnose the USM in use and operation.

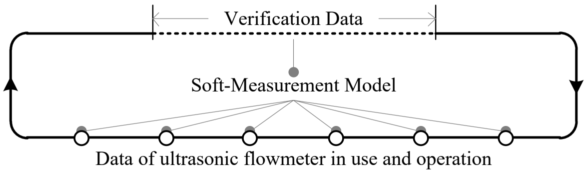


Figure 1. Schematic of the application of supervised learning in ultrasonic flowmeters inspection in-use method

2. USM and supervised learning algorithms

Currently, the USM used in natural gas measurement are facing the challenges of traceability and reproducibility with the gradual integration of digital technologies. Through the methods such as machine learning or digital twin, it is possible to ensure the consistency and reliability of the measurement during the use of the instrument, as well as the uncertainty of the measurement system.

2.1 Multipath ultrasonic flowmeter

The working principle of the USM is shown in Figure 2, the ultrasonic signal is opposite between the two ultrasonic transducers, and when the gas in the pipeline is stationary, the ultrasonic propagation time is the same in both directions, if the fluid flows, the gas velocity has a component in the direction of the sound route, so that there is a difference between in the upstream and downstream propagation time of the ultrasonic signal.

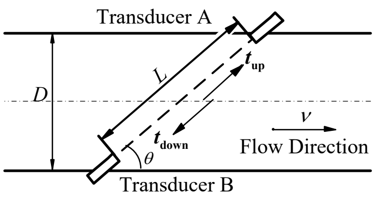


Figure 2: Principle of ultrasonic flowmeter

the upstream and downstream propagation time of the ultrasonic can be described as

$$\begin{cases} t_{up} = \dfrac{L}{c - v\cdot\cos\theta} \\ t_{down} = \dfrac{L}{c + v\cdot\cos\theta} \end{cases} \tag{1}$$

where v is the line average flow velocity on the sound path of the ultrasonic flowmeter, and c is the speed of sound on the sound path.

Multipath ultrasonic flowmeter is to arrange multiple paths in the pipeline according to a certain rule, so that each sound path is distributed in different positions of the gas in the pipeline, which can reflect the flow conditions of each position respectively. A cross 4-paths ultrasonic flowmeter is selected for the experiment, the structure schematic of which is shown in figure 3.

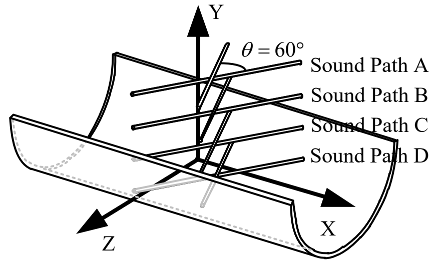


Figure 3: Structure of cross 4-paths ultrasonic flowmeter

To obtain an accurate estimation of the flow rate over the cross-section of the pipe, the weighted integration and correction of the flow velocities on the individual sound paths is carried out, which can be described as

$$Q_v = A\cdot K\cdot V = A\cdot K\cdot \sum_{i=1}^{4} w_i v_i \tag{2}$$

where A is the cross-sectional area of pipeline, Q is the measured flow rate of the USM, K is the flow velocity correction factor, v_i is the flow velocity on the i^{th} sound path, and w_i is the integration weight related to the i^{th} sound path. From equation (1) and (2), it can be seen that the SoS has a certain mathematical relationship with measurement flow rate, but since the K is private, the relationship is unknown, the SoS cannot provide a comprehensive diagnosis and monitor for instruments. In order to form a more complete in-use inspection process of USM, it is necessary to monitor all parameter indicator that affect the accuracy of it, and clarify the meaning of each, establish a flow rate deviation estimation model, and ensure the accuracy of flowmeter in use.

2.2 Flow rate deviation estimation algorithms

Supervised learning algorithms is a kind of machine learning method commonly used in regression problems. Since the measurement performance of the USM will change continuously over time, this paper uses the supervised learning method to train and the estimation model with the data from the verification of the flowmeter, which can continuously accumulate "learning" experience with the measurement data, so as to judge whether the performance meets the accuracy requirements.

(a). RF algorithm

The Random Forest (RF) algorithm is a learner that is integrated by multiple decision trees. It has the advantages of high estimation accuracy and controllable generalization error. It can effectively avoid the phenomenon of "overfitting" and is suitable for high-dimensional eigenvector space, these factors provide the feasibility of applying random forest to the in-use verification method of USM.

(b). BP-ANN algorithm
Back Propagation-Artificial Neural Network (BP-ANN) as a typical algorithm in supervised learning, it has strong self-learning and generalization ability, and its unique reverse error propagation method also ensures the accuracy of learning. The multi-layer neural network has good characterization ability for many complex practical problems, and has a strong nonlinear mapping ability in dealing with multi-feature high-dimensional data. It can respond quickly to high-dimensional variables that affect the accuracy of USM.

Digital metrology is used for the in-use inspection of ultrasonic flowmeter, and Random Forest and BP-ANN are used to construct the flow rate deviation estimation model based on the supervised learning algorithms [5].

3. Data acquisition and pre-processing

The tests for a DN200 USM with cross 4-paths were conducted in the high-pressure close loop gas flow standard facility in National Institute of Metrology, China. With 4 sets of turbine meters as its master meters, the facility can operate at the maximum flowrate of 1400 m^3/h under 2.5 MPa pressure, and the expanded uncertainty was 0.21% (k=2). The facility fully meets the verification test requirements of ultrasonic flowmeter.

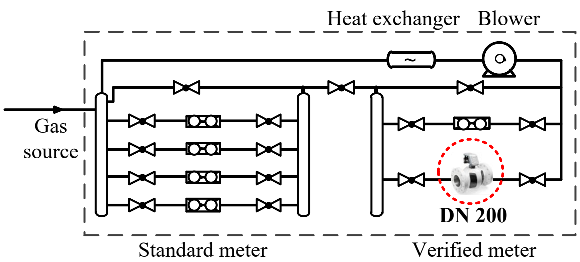


Figure 4. The close loop gas flow standard facility in NIM

3.1 Design of test experiment
This article is based on the setting of the USM verification test flow rate point in AGA No.9 Report, and combining with the flow rate range of the facility system, the maximum flow rate when tested is set 1200 m^3/h. The test points are set to six different flow rate, 0.70 Q_{max}, 0.50 Q_{max}, 0.40 Q_{max}, 0.30 Q_{max}, 0.20 Q_{max} and 0.10 Q_{max} (Q_{min}) respectively. Two groups of test experiments with an interval of one month were carried out respectively under similar pressure and temperature. The first set of test data is used as verification samples for training data of the flow rate deviation estimate model of the ultrasonic flowmeter, and the second set of data is used as inspection samples for model checking.

.

3.2 Parameters filtering and feature extraction
Four types of parameters can be artificially extracted through analysis based on the data obtained from the test experiments, which are signal quality parameter, flow field state parameter, speed of sound checking parameter, and measurement performance parameter.

a. signal quality parameter
The signal automatic gain control value (*S-AGC*) and the signal-noise ratio (*SNR*) can intuitively reflect the interference degree of the ultrasonic signal in the flow field. The operation status of the flowmeter can be effectively monitored by observing them.

b. flow field state parameter
The flow field operation of the flowmeter can be simply analysed through the average flow velocity (V) and flow velocity of each sound path (v_i), and the flow rate (Q_v) can reflect the working condition of the flowmeter.

c. speed of sound checking parameter
The speed of sound each path (c_i) can indirectly reflect the operation of the flowmeter. At the same time, the theoretical speed of sound (C) can be calculated by the equation from AGA No.10 Report according to the temperature, pressure and composition data of the working conditions, and the speed of sound deviation can be calculated by using the theoretical value and the measure the average speed of sound (c). The equation for calculating the speed of sound deviation is

$$E_c = \frac{c - C}{C} \times 100\% \tag{3}$$

d. measurement performance parameter
This paper uses the measure flow rate deviation as the basis for evaluating the measurement performance of the ultrasonic flowmeter, which is calculated from the measured value of ultrasonic flowmeter and the actual flow rate measured (Q_s) by the standard facility. The calculation equation is

$$E_Q = \frac{Q_v - Q_s}{Q_s} \times 100\% \tag{4}$$

A total of 15 indicators can be extracted from the parameters (a), (b) and (c), the indicator set is not dimensionally reduced in order to understand and observe the changes of the data more comprehensively, all of them are used as the features input for estimate model, and parameter (d) as the output of it.

3.3 Data Analysis and Processing
Stable working conditions are the premise to ensure the accurate measurement of the flowmeter, factors such as gas temperature, pressure and environmental conditions will affect the test results during the test. In this experiment, the maximum change of the temperature of each group of tests was 0.31 °C, and the maximum change range of the pressure was 1.76 kPa, which accounted for no more than 0.07 %, which fully meet the requirements of USM. Meanwhile, the repeatability is used to characterize the stability of the measurement results, and its calculation formula is as equation

$$r=\sqrt{\frac{\sum_1^n\left(E_{Qi}-\overline{E}_Q\right)^2}{n-1}} \tag{5}$$

where $\overline{E}_Q$ is average value of flow rate deviation, E_{Qi} is instantaneous value of flow rate deviation, and n is the number of test data. The test results show in Table 1, the repeatability of the measured flow rate deviation of the test data at different test points is no more than 0.05 %, which also fully meets the measurement requirements of ultrasonic flowmeter.

Table 1: Test data results.

Test point	Verification samples		Inspection samples	
(m^3/h)	E_Q (%)	r (%)	E_Q (%)	r (%)
1120	0.25	0.01	0.23	0.04
800	0.21	0.01	0.25	0.04
640	0.15	0.01	0.18	0.02
480	0.15	0.01	0.19	0.05
320	0.10	0.02	0.09	0.01
160	0.10	0.02	0.07	0.03

Other factors such as environmental interference during the test will also affect the experimental results. The box plot can reflect the distribution and fluctuation of the data. The data with 1120 m^3/h flow rate points in the verification samples is a 300×16 matrix where each of column contains 15 input indicators and 1 output indicator, and it is analysed an example. The boxplot of the data is shown in Figure 5.

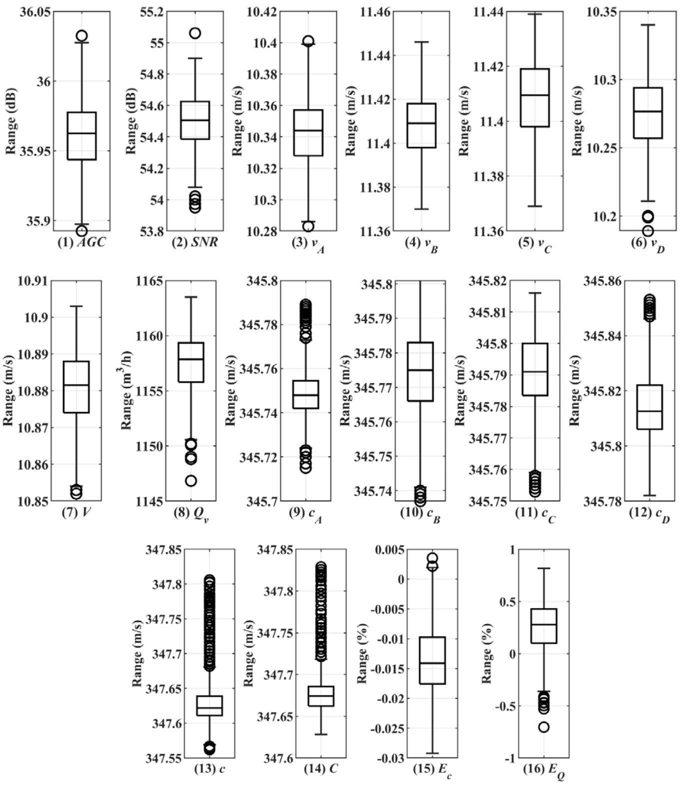


Figure 5. Box plot of data

The results show that the stability of each path value, measured average value and the theoretical of the speed of sound is not good, and its fluctuation is relatively larger than other indicators, which also shows that there is a certain gap between the data distribution and the normal distribution. Although the above proves that the measurement result meet the accuracy requirements, its box plot shows that there are still a small number of outliers.

The data are analysed and denoised by the Fast Fourier Transform in order to further ensure the validity of the data and eliminate the interference of other factors [6]. The flow rate deviation data in the verification sample as an example for analysis are taken, and the frequency is calculated according to the data sampling interval of 2 seconds/time and the number of data, all the frequency features of the data are obtained.

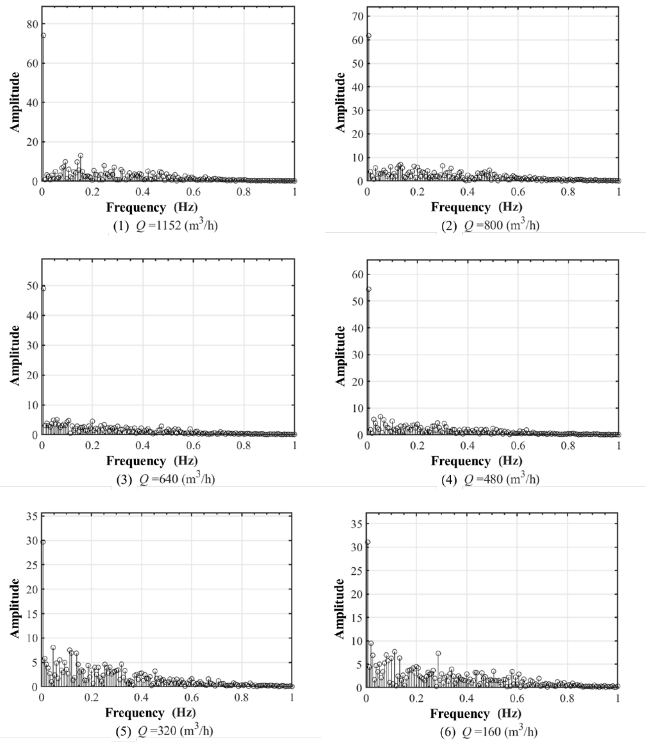


Figure 6. FFT spectrogram of flow rate deviation data

The results show that there is a significant frequency band component in the data near the low frequency. The amplitude of the low frequency band gradually decreases as the test flow rate value decreases, and the signal amplitude of the high frequency band increases. The spectral features distribution of the data under different flow rate values is similar, and the amplitude of the main frequency is significantly different from that of the other frequencies. It can be considered that the information in the relatively high-frequency part of the data is system noise caused by interference factors. In this regard, the Gaussian window function is used to convolve the data, and then the time domain signal of the data is restored by

Inverse-Fast Fourier Transform to realize the noise reduction of the data.

Based on the number of test samples, the length of the Gaussian window function is set to 151, which is used to scan each group of data separately. The result of the data after denoising is shown in Figure 7.

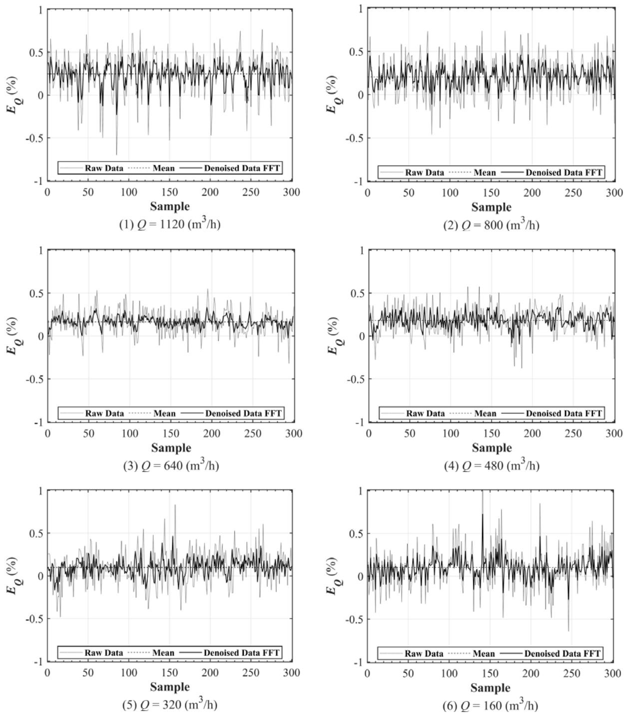


Figure 7. Noise reduction of flow rate deviation data

The results show that the original data has obvious steep rise and drop, noise reduction effectively reduces the fluctuation of data compared with before, and the amplitude of the data in the time domain is reduced. Therefore, it is believed that this method can filter out the noise introduced by the interference factors to the data to a certain extent, and the denoised data can better represent to and reflect the actual working conditions. Next, denoise all group data based on this method.

3.4 Features scaling of data

Excessive numerical value differences among various data such as speed of sound, flow velocity, and flow rate will lead to the dominance of indicators with large magnitudes, making them unable to be directly used for calculation. This paper adopts the features scaling method to reduce the deviation caused by the large numerical span of different dimensions, and eliminate the difference in the weight of each indicator caused by different.

4. Results discussion and model uncertainty

4.1 Estimate model building

Each flow rate point has about 300 and 100 pieces of data in the verification and inspection samples respectively, all of which are continuous data. Taking the verification samples as the training set and the inspection samples as the test set, the data is trained and estimated by two algorithms, and the performance measurement and comparison of the results are carried out.

(a). RF model

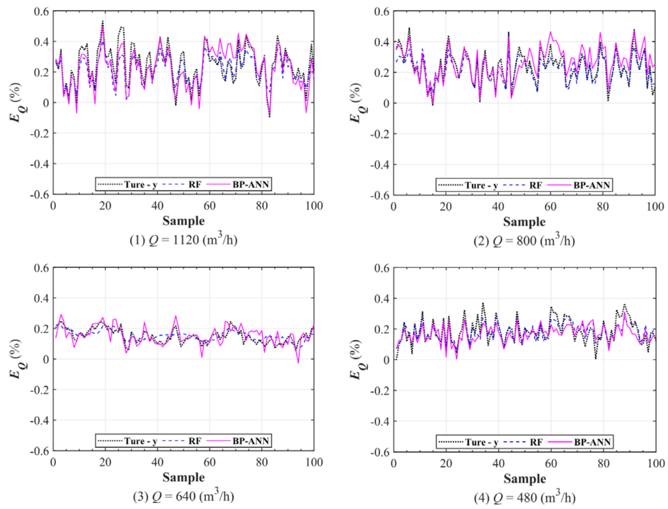
Each decision tree in the random forest participates in the learning process and needs to grow as much as possible without pruning. In this paper, the number of forest trees is set to 160, and the number of features used by each decision tree is set to account for 25 % of its total. Using the bootstrap aggregating method, the samples were sampled from the 6 groups of training data with replacement, and about 36.8 % of the data did not participate in the fitting of the training set model, but were used to calculate the out-of-bag error. The results are used as measure of the Random Forest estimate model.

(b). BP-ANN model

The BP-ANN model is mainly divided into input layer, hidden layer and output layer. The neurons in the hidden layer use the weighting and activation function to calculate the indicators by the input layer, and then the result produces by output layer. In the modelling process, the number of layers of input layer, hidden layer and output layer are all set to 1. The 15 indicators are input as a 15-row numerical matrix, the number of neurons in the hidden layer is 9, and the activation functions of the hidden layer and the output layer are sigmod and linear function respectively. The results are used as measure of the BP-ANN estimate model.

4.2 Estimated performance comparison

By setting the optimal parameters of the model, the Random Forest and BP-ANN flow rate deviation estimate models are established respectively, and 100 data samples in the test data set are used to estimate and compare the results, which is shown in Figure 8.



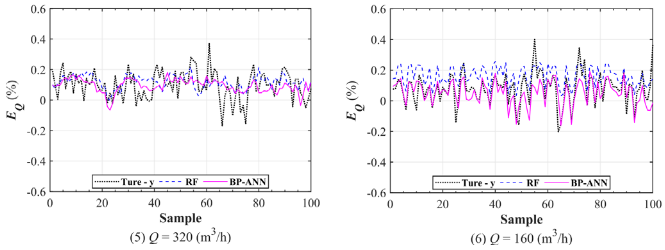


Figure 8. Evaluation result of models

In Figure 8, two lines RF and BP-ANN are the estimated values of the flow rate deviation by the corresponding regression model, the True-y is the actual measured value of the flow deviation. In Figure8. (1), (2) and (4)，it shows that the estimated results of flow deviation are basically consistent with the actual values, and to reflect its characteristics, the distribution trend of the data also has a good consistency. In Figure 8. (3) and (5), it shows that the robustness of the RF model is poor, the estimated values does not match well with the True-y. In the Figure 8. (6), the fitting degree of the two models is poor compared to other flow points, but the overall trend is consistent with the True-y.

In order to further compare the prediction performance of the two models, two regression model evaluation criteria are selected, which are: Coefficient of Determination (R^2) and Root Mean Squared Error (E_{RMSE}) for comparison. The results are shown in Table 2.

Table 2: Model Estimation Results Evaluation.

Test point (m³/h)	***R²***		***ERMSE* (%)**	
	RF	BP-ANN	RF	BP-ANN
1120	0.86	0.86	0.06	0.07
800	0.92	0.84	0.05	0.07
640	0.55	0.71	0.05	0.04
480	0.85	0.73	0.04	0.05
320	0.46	0.42	0.08	0.08
160	0.73	0.65	0.11	0.09
Mean	0.72	0.70	0.07	0.07

The range of R^2 of the RF model is 0.46 ~ 0.91, and the BP-ANN model is 0.42 ~ 0.88, in contrast, the BP-ANN model has a smaller value of R^2, from the overall situation, the mean of the RF model R^2 is relatively higher, but the order of magnitude difference between the two is not large. Observe the E_{RMSE} of the two estimation models at different flow rate points at the same time. The range of the values for BP-ANN model is small compared to RF. From the overall situation, the mean of the two models E_{RMSE} is basically the same. The results show that it is feasible to apply the supervised learning method to the inspection of ultrasonic flowmeters in-use. It can effectively estimate the deviation of flow rate, so as to monitor and diagnose the performance of USM in use and during operation.

4.3 Uncertainty assessment of model

In this paper, the Monte Carlo propagation probability distribution method (MCM) is proposed to assess the uncertainty introduced by the estimation model of RF and BP-ANN, which is suitable for the situation that the model is nonlinear or complex, and the probability density function (PDF) of the input features are asymmetric distribution. MCM is through calculate discrete sampling of the output PDF by discretely sampling the input, and then propagates the distribution of the input through model to obtain the best estimate of the output.

KL divergence (Kullback Leibler, KL), called relative entropy, is an asymmetric measure to compare the difference between two PDFs, by calculating the entropy of the probability distribution to determine how much information is lost by its best approximate distribution function, its calculation formula is

$$D_{KL}\left(p\|q\right)=\sum_{i=1}^{n}p(x_i)\left(\log\frac{p(x_i)}{q(x_i)}\right) \tag{6}$$

This paper finds the best approximate distribution of the PDF of each input by calculating the value of KL. The results are shown in Table 3.

Table 3: Approximate distribution of input.

Index	Indicator	Approximate distribution	KL
1	AGC	Gaussian	0.03
2	SNR	Gaussian	0.03
3	v_A	Gaussian	0.03
4	v_B	Gaussian	0.03
5	v_C	Gaussian	0.11
6	v_D	Gaussian	0.11
7	V	Gaussian	0.07
8	Q_v	Gaussian	0.19
9	v_A	t	0.29
10	v_B	t	0.29
11	v_C	t	0.24
12	v_D	t	0.22
13	c	t	0.22
14	C	Gaussian	0.09
15	E_c	Gaussian	0.05

The results show that the best approximation distribution for PDFs for most of the input is the normal distribution, and the input related to the speed of sound is the t-distribution. This method is used to determine the function of approximate distribution of the input data of other flow rate points.

Next, the experimental sample size (M) of each indicator in MCM and the probability (p = 95 %) is set to calculate the coverage interval. The M should be much larger than $1/(1\text{-}p)$ in theory, the best suitable size of sample is set to 1×10^5 by independent test. The generated MCM samples were imported into the RF and BP-ANN models respectively, and 1×10^5 measured flow rate deviation can be obtained for each flow rate point. The mean of the estimated value ($\overline{E}_Q$) and the standard deviation value $u(y)$ from the three MCM outputs and the value with a coverage interval of 95 % were taken as the final result.

Table 4: MCM results of RF model ($M=10^5$),

Test point (m^3/h)	$\overline{E}_Q$ (%)	$u(y)$ (%)	Coverage interval Interval (%)	Width (%)
1120	0.23	0.12	[0.01, 0.43]	0.42
800	0.21	0.10	[0.04, 0.41]	0.37
640	0.15	0.02	[0.11, 0.18]	0.07
480	0.18	0.06	[0.06, 0.30]	0.24
320	0.07	0.02	[0.04, 0.11]	0.07
160	0.11	0.09	[-0.04, 0.26]	0.30

Table 5: MCM results of BP-ANN model ($M=10^5$).

Test point (m^3/h)	$\overline{E}_Q$ (%)	$u(y)$ (%)	Coverage interval Interval (%)	Width (%)
1120	0.17	0.28	[-0.32, 0.60]	0.92
800	0.21	0.20	[-0.09, 0.55]	0.66
640	0.11	0.21	[-0.26, 0.43]	0.69
480	0.18	0.14	[-0.05, 0.41]	0.46
320	0.09	0.27	[-0.34, 0.53]	0.87
160	0.08	0.27	[-0.37, 0.50]	0.87

The results in Tables 4 and 5 show that the estimated result of the deviation of the RF model at each flow rate point is the range of 0.07 % ~ 0.23 %, and the result of BP model is in the range of 0.08 % ~ 0.21 %, the both are consistent with the accuracy requirements of flowmeter. In addition, the mean of uncertainty of RF model is better than that of BP-ANN, and the maximum change range of the uncertainty of results of the RF and BP-ANN under each flow rate is no more than 0.10 % and 0.14 % respectively. In contrast, the uncertainty of the RF estimation model is significantly better than that of the BP-ANN, so it can be considered that the in-use inspection of the ultrasonic flowmeter based on the RF algorithm is a better choice.

5. Conclusion

In this paper, the supervised learning method and the digital metrology technology are combined to establish a soft-measurement model of the USM, and the MCM is used to assess the uncertainty of it. The results show that the model can effectively estimate the flow rate deviation of flowmeter, so as to monitor and diagnose the performance during operation. Thus, a new reference is provided for the inspection method of USM in use.

References

[1] AGA Report No.10: *Speed of Sound in Natural Gas and Others Related Hydrocarbon Gases*[S]. 2003.

[2] ISO 20765-1: *Natural gas - Calculation of Thermodynamic Properties -Part 1: Gas Phase Properties for Transmission and Distribution Applications*[S]. 2005.

[3] PTB Technische Richtlinien G-18: *Messgeräte für Gas - Anforderungen an Dauerreihenschaltung vo n zwei Ultraschallgaszählern*[S]. 2012.

[4] Wu Y, Song CF, el. Practice of Intelligent Quality Control Technology in Verification of Natural Gas Flowmeter[J]. Digitization and Intelligentization, 40(06):637-642. 2021

[5] Wu E. *Machine Learning Yearning*[M]. Andrew, NG, 2018.

[6] Cooley JW, Tukey JW. An Algorithm for the Machine Calculation of Complex Fourier Series[J]. Mathematics of computation, 19(90):297-301, 1965.